



研究与开发

基于相位谱的翻录语音攻击检测算法

李璨, 王让定, 严迪群, 陈亚楠

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 因与原始语音具有高度相似性, 经高保真设备回放的翻录语音常被不法分子用于对说话人认证(ASV)系统进行攻击, 以达到非法认证的目的。为提高系统抵抗翻录语音攻击的顽健性, 通过研究原始语音与翻录语音产生的实际过程, 发现两者在频率域相位上有明显差异, 并在此基础上提出了一种基于相位谱的翻录语音检测方法。分析讨论了 FFT 和不同偷录、回放设备对翻录语音检测率的影响。实验结果表明, 该方法能够准确地判断待测语音是否为翻录语音, 其检测率达到了 99.04%。并且, 将该算法加载到说话人识别系统中, 使系统的等错误概率(EER)降低了约 22%, 有效提高了系统抵抗翻录语音攻击的性能。

关键词: 说话人认证系统; 翻录语音检测; 相位谱

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017126

Recapture voice replay detection based on phase spectrum

LI Can, WANG Rangding, YAN Diqun, CHEN Yanan

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Due to a high similarity between the recaptured voice recorded by high-fidelity ripping equipment and the original voice, the automatic speaker verification (ASV) system used to be attacked illegally by the recaptured voice. In order to improve the ability of resisting the attack, a recaptured voice detection method was proposed based on the difference of phase spectrum between original and recaptured voices for the ASV system. In addition, the effects of different recording and replay devices, the FFT were discussed. Experimental results show that the proposed method can accurately recognize the recording voice, of which detection rate is 99.04%. Meanwhile, the equal error rate (EER) of the ASV system has dropped about 22% with this method being integrated, which indicates that the system's ability of resisting playback attack is enhanced.

Key words: ASV system, recaptured voice detection, phase spectrum

收稿日期: 2017-01-23; 修回日期: 2017-03-20

通讯作者: 王让定, wangrangding@nbu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61672302, No.61300055); 浙江省自然科学基金资助项目 (No.LZ15F020010, No.Y17F020051); 宁波大学科研基金资助项目 (No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503); 宁波大学王宽诚幸福基金资助项目

Foundation Items: The National Natural Science Foundation of China (No. 61672302, No.61300055), Natural Science Foundation of Zhejiang Province of China (No.LZ15F020010, No.Y17F020051), The Scientific Research Foundation of Ningbo University (No.XKXL1405, No.XKXL1420, No.XKXL1509, No. XKXL1503), K.C. Wong Magna Fund in Ningbo University



1 引言

在信息时代,录音设备(如录音笔、智能手机等)日趋低价和高保真,回放设备也越来越精准,为录音造假提供了可能。翻录语音经录音设备偷录、回放设备回放,与原始语音难辨真假,一些说话人认证系统也无法辨别。而且翻录语音因偷录设备体积小、易偷录、成功率高等优势,已成为攻击语音认证系统中最易实施的方法。因此,对翻录语音检测十分迫切且具有现实意义。

近年来,国内外研究学者对翻录语音检测展开研究,并取得了一定的研究成果。其中检测算法主要分为以下两类:一是基于语音产生随机性的检测算法。Shang W 等人^[1,2]根据语音的随机性,提出了一种基于 Peak map 特性的录音回放检测算法,该算法通过计算原始语音与翻录语音 Peak map 特性的相似度,判断待测语音是否为翻录语音。若相似度大于设定的阈值,判定为翻录语音;反之,判定为原始语音。在此基础上,Jakub G 等人^[3]对该算法进行了改进,在 Peak map 特性中加入了各频率点的位置关系,Wu Z 等人^[4]加入光谱峰作为检测特征,但该算法只能针对文本相关的识别系统,且只针对一种偷录设备,存在较大的局限性。二是基于语音信道的检测算法。张利鹏等人^[5]根据信道模式特征,用语音数据的静音段对信道建模,检测待测语音与训练语音的信道是否相同,从而判断是否为回放攻击。王志锋等人^[6,7]根据原始语音与翻录语音产生的信道不同,提取信道模式噪声,利用 SVM(support vector machine,支持向量机)得到了很好的分类结果。Villalba 等人^[8,9]依据远距离的录音会受到噪声和混响的影响,提出了针对远距离偷录语音的检测方法。Chen Y N 等人^[10]根据设备信道对语音编码过程的影响,提出基于长窗比例因子的回放语音检测算法。但该类方法提取的信道模式噪声并非准确,且录制语音的设备过于单一,且未对多种不同的偷录

设备及回放设备进行分析与研究。

伴随着信息化的快速发展,信息安全形势愈加严峻^[11,12]。然而,目前针对多种不同偷录设备与回放设备的翻录语音检测的研究关注较少。在现实生活中,各种高质量低价格的高保真录音设备随处可见,如录音笔及各种智能手机。这类偷录设备携带便利且不易察觉,且获得的翻录语音与原始语音相似性较高,因此这种录音设备是目前较为主流的偷录设备。同时,不同的回放设备对翻录语音的产生也有较大的影响。但是,由于偷录设备和回放设备种类繁多,不同的录音设备中传感器和信号采集电路存在差异,产生不同的设备信息。这无疑增加了对不同录音设备的翻录语音攻击检测的难度。

本文通过分析研究原始语音与翻录语音产生的实际过程,发现翻录语音经历了一次偷录与回放过程后,其相位会发生不同程度的变化,根据相位谱的不同,提出了一种基于相位谱的翻录语音检测算法。为提高算法的实用性,弥补现有算法涉及录制设备单一的问题,本文讨论了不同偷录设备和回放设备对检测率的影响,并对影响程度作了分析。实验结果表明,将本文检测方法加载到目前主流的 GMM-UBM 和 i-vector 说话人识别系统中后,系统抵抗翻录攻击的性能得到了极大的提高。

2 原始语音和翻录语音产生的实际过程

原始语音和翻录语音产生的实际过程如图 1 所示。从说话人处直接得到的语音本文称为原始语音,也称为合法语音。将说话人语音用偷录设备进行偷录,回放设备进行回放得到的语音称为翻录语音。

由图 1 可知,翻录语音比原始语音多经历了一次高保真音响系统。高保真音响系统原汁原味地还原了原始语音,使得原始语音与翻录语音具有高相似度,但该系统不可避免地对语音信号进

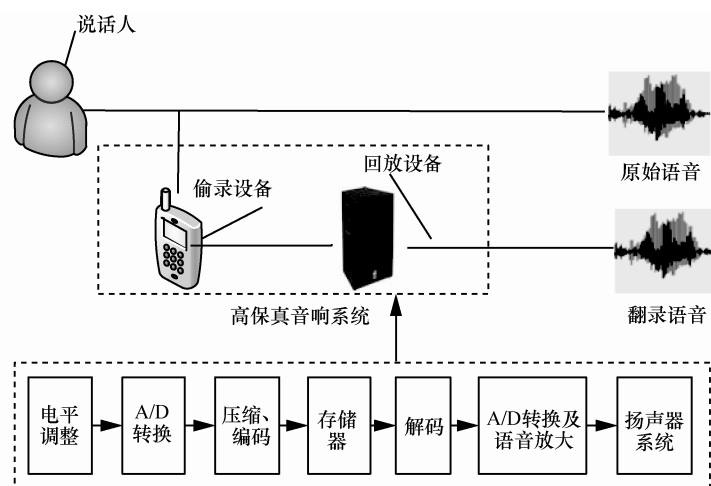


图1 原始语音和翻录语音产生的实际过程

行电平调整、A/D 转换、编解码等一系列的操作，使得翻录语音与原始语音还是存在着一定的差异。

原始语音和翻录语音的波形如图 2 所示，其中图 2 (a) 是一段经 Aigo R6620 录音笔原始录制的语音信号的波形图，该语音信号的具体内容为普通话朗读的“开窗 关灯 亮度佳 播放音乐”。图 2(b)、图 2 (c)、图 2 (d) 是对应的翻录语音信号的波形，翻录过程中的偷录设备分别为 iPhone6、Mi4 和 Sony PX440，回放设备则都选择的是 Huawei AM08。从图 2 可以看出，翻录语音信号较原始语音还是产生了一些失真，只是针对不同的翻录设备，失真的大小略有区别。例如，在图 2 所示的 3 款偷录设备中，iPhone6 对应的语音信号失真最小，Sony PX440 对应的语音信号失真最大。

图 3、图 4 的波形分别与图 2 一一对应，其回放设备分别为 Philips DTM3115 和 Yamaha

TSX-140。由图 2、图 3、图 4 可以看出，无论是偷录还是回放设备，对翻录语音都有一定的影响，但对翻录语音有何影响且影响程度如何，需要进一步探索。同时，在时域上很难区分原始语音与翻录语音，但翻录语音经过高保真系统的电平调整、A/D 转换、编解码等一系列的操作，该系统将对不同频率信号相位产生超前或者滞后的影响，会使得翻录语音的相位产生较大程度的失真。

3 特征提取和翻录语音攻击检测

相位谱检测算法主要由语音预处理、相位谱特征提取、特征选择、分类识别 4 个部分组成。在特征选择上，本文将相位统计平均作为检测特征，并运用 SVM-RFE 算法进行特征筛选，采用 LIBSVM 分类器进行分类识别。为验证该算法的

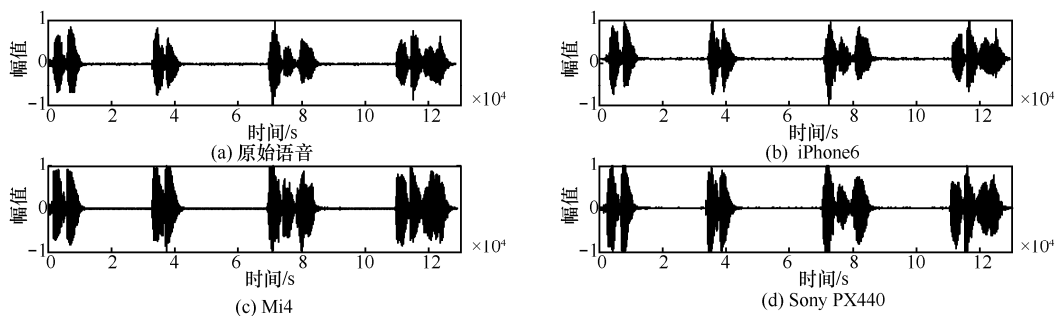


图2 原始语音和翻录语音的波形 (Huawei AM08)

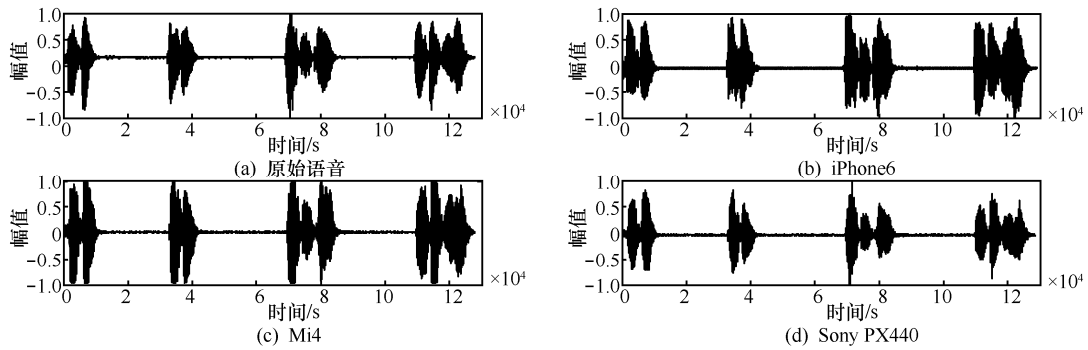


图3 原始语音和翻录语音的波形 (Philips DTM3115)

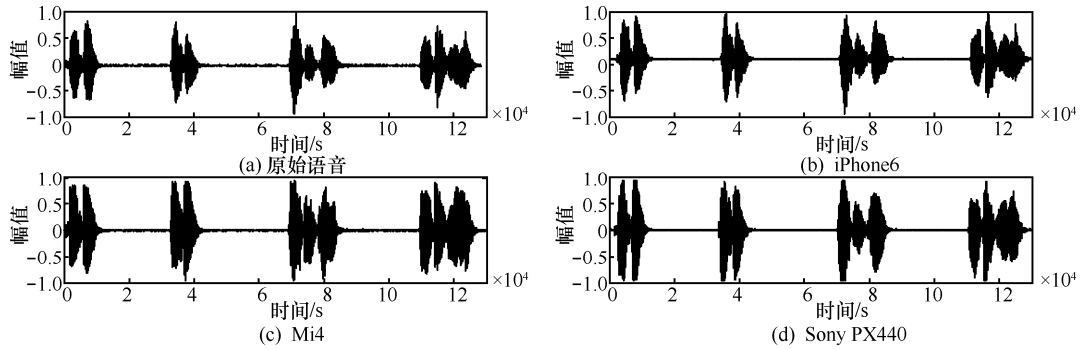


图4 原始语音和翻录语音的波形 (Yamaha TSX-140)

有效性, 本文将该检测算法加载到了说话人识别系统中, 并对该系统防御翻录语音的攻击能力进行了检测。

3.1 特征提取

相位信息对感知有着不可忽视的作用, 相位谱反映信号方向随频率变化的规律, 含有大量的信息^[13]。本文算法首先根据待测语音提取相应的相位。设待检测语音信号为 x , 将语音信号分帧处理后, 对其进行 FFT, 得到:

$$x_i(k) = \sum_{n=0}^{N-1} x_i[n] e^{-j \frac{2\pi}{N} kn} \quad (1)$$

其中, $k=0,1,2,\dots, N-1$; $i=1,2,\dots,S$, S 表示总帧数。

对于语音信号第 i 帧, 求其相位 $\phi_i(k)$:

$$\phi_i(k) = \tan^{-1} \left(\frac{-\sum_{n=0}^{N-1} x_i[n] \sin \frac{2\pi}{N} kn}{\sum_{n=0}^{N-1} x_i[n] \cos \frac{2\pi}{N} kn} \right) \quad (2)$$

为表征相位谱的变化程度, 本文选取均值统计特性。对于第 i 帧第 j 频率点相位 $\phi_{i,j}(k)$, 求其幅值:

$$\|\phi_{i,j}(k)\| = \phi_{i,j}(k) \times \overline{\phi_{i,j}(k)} \quad (3)$$

然后求其第 j 频率点相位的统计平均 $\varphi_j(k)$:

$$\varphi_j(k) = \frac{1}{S} \sum_{i=1}^S \|\phi_{i,j}(k)\| \quad (4)$$

最后, 对 $\varphi_j(k)$ 进行归一化处理, 得到语音信号相位谱特征。初步分析可知, 相较于原始语音, 翻录语音经历了一次高保真系统, 该系统会对不同频率信号相位产生超前或滞后的影响。另外, 受录音设备的影响, 语音在翻录过程中将会引入一定的设备噪声, 这种噪声是录音设备所固有的, 会使语音的相位谱产生明显的失真。

图5、图6、图7分别是3种不同的回放和偷录设备的原始语音与翻录语音的相位谱特征图。图中 FFT 采样点数 N 为 1 024, 根据实序列离散

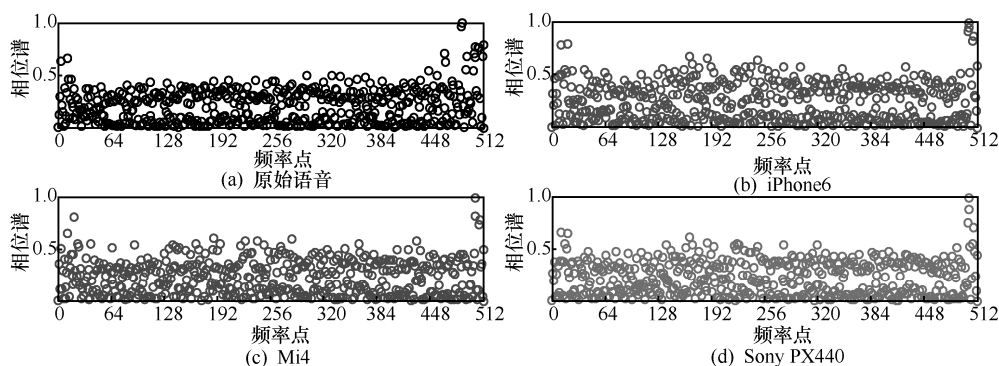


图5 原始语音与翻录语音相位谱特征 (Huawei AM08)

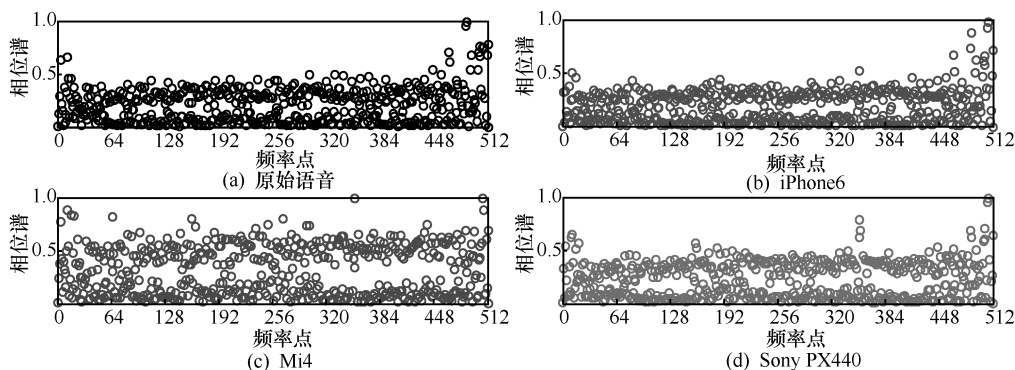


图6 原始语音与翻录语音相位谱特征 (Philips DTM3115)

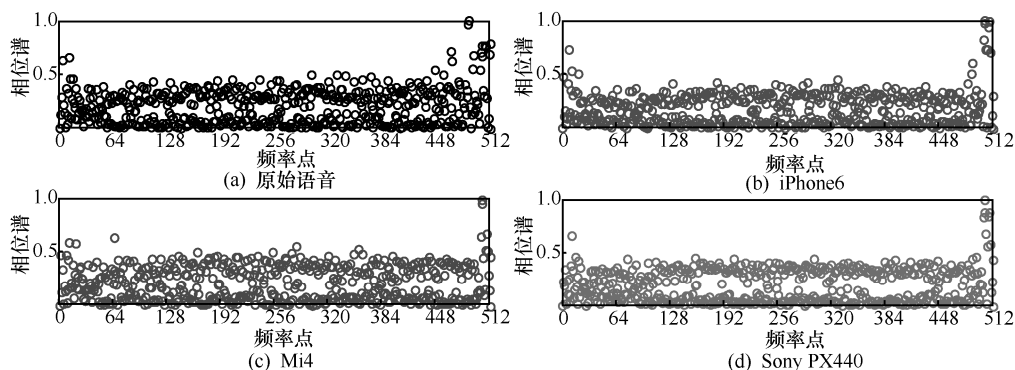


图7 原始语音与翻录语音相位谱特征 (Yamaha TSX-140)

傅里叶变换的共轭对称性, 本文选取前 512 个值。其中图 5 (a)、图 6 (a)、图 7 (a) 表示一段由 Aigo R6620 录音笔录制的语音内容为“开窗 关灯 亮度佳 播放音乐”的原始语音的相位谱, 图 5 (b)、图 5 (c)、图 5 (d)、图 6 (b)、图 6 (c)、图 6 (d)、图 7 (b)、图 7 (c)、图 7 (d) 分别是偷录设备为 iPhone6、Mi4 和 Sony PX440, 回放设备为 Huawei AM08、Philips DTM3115 和 Yamaha TSX-140 的翻录语音相位谱。

由于偷录设备和回放设备均是翻录过程的重要设备, 它们均会在翻录语音中留下“痕迹”。当回放设备相同时, 图 5、图 6、图 7 中的 (a) 与 (b)、(c)、(d) 在高频区有明显的不同, (b)、(c)、(d) 在高频区相似度较高。当偷录设备相同时, 不同的回放设备其相位谱区分较为明显。但是总体来看, 针对不同的偷录设备和回放设备的翻录语音, 其相位信息与原始语音具有较大的失真。为了实现更好的性能测试, 提高运算效率, 选择



有效的特征维数, 本文采用 SVM-RFE 算法进行特征选择。

SVM-RFE 算法^[14]的输出结果是按照特征的重要性对特征进行排序的列表。本文将上述所提取的相位特征进行特征筛选。这里使用的 FFT 采样点数为 1 024, 样本数目为 14 000 (原始语音 1 400, 翻录语音 12 600)。其中, *RANK* 为最终的排列等级, *AVG_RANK* 为 5 次交叉验证 *RANK* 结果的平均值, 特征索引 *N* 为前 512 维特征。SVM-RFE 对特征排序的结果 (前 20 列) 见表 1。

表 1 SVM-RFE 特征排序的结果

<i>RANK</i>	特征索引 <i>N</i>	<i>AVG_RANK</i>	<i>RANK</i>	特征索引 <i>N</i>	<i>AVG_RANK</i>
1	509	1	11	132	22.2
2	354	2.4	12	508	23.4
3	3	6.2	13	353	24
4	21	12	14	505	24.4
5	507	13.6	15	200	30.4
6	355	15.4	16	84	30.6
7	186	16.8	17	506	33
8	501	18	18	258	35.2
9	512	18.8	19	498	35.6
10	2	22.2	20	196	38

由表 1 可知, 在前 20 列特征重要性的排序中, 特征索引 *N* 为 509 的特征翻录语音失真最大。整体来看, 特征索引 *N* 较大的值所占的比例较高, 即翻录语音较原始语音在高频区失真较为明显。因此, 为提高检测效率, 选择 $RANK \geq 10$ 的 10 维特征作为最终的有效特征。

3.2 翻录语音攻击检测

为验证基于相位谱的翻录语音检测算法的有

效性, 本文将该算法加载到了说话人识别系统中, 如图 8 所示。

具体的操作步骤如下: 语音经采集设备采集, 同时输入说话人识别系统和翻录语音检测模块中; 在翻录语音检测模块, 提取语音信号的相位谱特征, 对语音进行分类识别。若输入的语音为翻录语音, 则判决 1 为 0; 否则, 判决 1 为 1; 说话人识别系统对输入语音进行判决, 若输入语音为说话人语音, 则判决 2 为 1; 否则, 判决 2 为 0; 在判决模块, 结合判决 1 和判决 2 的结果, 按表 2 规则对输入语音进行判断, 输出最终结果。

表 2 判决规则

判决 1	判决 2	结果
0	0	0
0	1	0
1	0	0
1	1	1

4 语音库

为了验证本文方法的有效性, 本文构建了一个符合研究目的的语音数据库。数据库的具体设置如下: 语料库来源于 863 语料库^[15]; 人员分布为: 18 男 14 女; 设备选取主要涉及语音采集设备、偷录设备、回放设备。设备的详细信息见表 3。

实验库的构建环境为安静房间, 阅读内容为语料库内容, 参与者依据自身朗读习惯用标准普通话进行录音, 并使用采集设备进行语音采集, 采集设备距参与者大约 20 cm, 本文将采集设备此

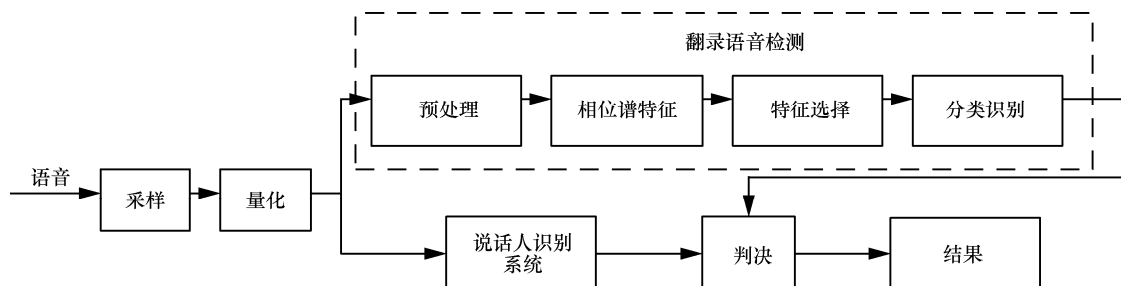


图 8 防翻录语音攻击的识别系统流程

表3 设备信息

	采集设备	偷录设备			回放设备		
类型	Aigo R6620	iPhone6	Mi4	Sony PX440	Huawei AM08	Philips DTM3115	Yamaha TSX-140
语音格式	wav	m4a	mp3	mp3	—	—	—
参数	16 kHz	44.1 kHz	44.1 kHz	44.1 kHz	—	—	—
	16 bit/s	64 kbit/s	128 kbit/s	192 kbit/s			

次采集到的语音称为原始语音。在参与者阅读以上语料的同时,将偷录设备同时打开到正常录制功能下,录制参与者的语音内容,根据实际情况,将偷录设备距离说话人大约 70 cm。在同样的环境下,将偷录设备采集到的语音经音响回放,并使用采集设备录制该回放语音。回放音响距离采集设备 20 cm 左右。将此次采集设备采集到的语音称为翻录语音。实验样本为 14 000 个(原始语音 1 400 个,翻录语音 12 600 个)。样本详情见表 4。

表4 原始语音和翻录语音样本详情

语音类型	回放设备	偷录设备	样本数量/个
原始语音	—	—	1 400
翻录语音	Huawei AM08	iPhone6	1 400
		Mi4	1 400
		Sony PX440	1 400
	Philips DTM3115	iPhone6	1 400
		Mi4	1 400
		Sony PX440	1 400
	Yamaha TSX-140	iPhone6	1 400
		Mi4	1 400
		Sony PX440	1 400

5 实验设置及结果分析

实验中使用 LIBSVM,它是一种监督是学习方法,广泛地应用于统计分类以及回归分析。分类过程如下:先根据第 3.1 节进行特征提取并使用 LIBSVM 进行训练得到模型,将待测语音在

LIBSVM 模型上进行测试并给出最终判断结果。本文将根据录音过程中的影响变量,分别分析其对检测翻录语音的影响。

5.1 FFT 变换点数的影响

相位谱反映信号方向随频率变化的规律,含有大量的信息。不同的 FFT 采样点数对相位信息产生较大的影响。随着采样点数的增加,语音相位谱所包含的信息越多,翻录语音经过高保真系统后,其相位失真越来越明显。为了寻找最佳的检测效果的采样点数,讨论了基于 FFT 采样点数为 128、256、512、1 024 时的翻录语音检测效果。实验采样 LIBSVM 分类器在 weka 平台上进行试验。实验样本为原始语音 1 400 个,翻录语音 12 600 个。检测效果见表 5,其中包括真正类率(true positive rate, TPR)、负正类率(false positive rate, FPR)、准确度(accuracy, ACC)等。

由实验结果可以看出,当 FFT 采样点数 N 为 1 024 时,对翻录语音有较好的检测效果,其检测率达到 99.04%。随着 FFT 采样点数的增大,包含的语音相位信息越多,越能较好地反映语音信号的本质特征。综合考虑,本文选取 1 024 作为最佳检测率时的 FFT 采样点数。

5.2 不同录音设备的影响

由第 5.1 节可知,当 FFT 采样点数为 1 024 时,对翻录语音检测效果最优。但不同的偷录设

表5 翻录语音检测率

FFT 采样点数	LIBSVM					
	TPR	FPR	ACC	召回率	精密率	F 值
128	92.20%	60.30%	92.21%	91.20%	92.20%	90.80%
256	95.80%	28.60%	95.78%	95.60%	95.80%	95.60%
512	98.10%	14.80%	98.14%	98.10%	98.10%	98.10%
1 024	99.00%	6.66%	99.04%	99.00%	99.00%	99.00%



表 6 不同偷录设备与回放设备的翻录语音检测

回放设备	偷录设备								
	iPhone6			Mi4			Sony PX440		
	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR	ACC
Huawei AM08	100%	0	100%	100%	0	100%	100%	0	100%
Philips DTM3115	99.80%	0.20%	99.78%	99.92%	0.10%	99.92%	99.80%	0.20%	99.78%
Yamaha TSX-140	100%	0	100%	99.71%	0.30%	99.71%	100%	0	100%

备与回放设备所含有的固有设备信息不同，它对相位谱的影响也就不同。因此，本节通过实验揭示不同偷录设备和不同回放设备对翻录语音检测率的影响。检测结果见表 6。

由表 6 可以看出，当回放设备为 Huawei AM08 时，来源于 3 种不同偷录设备的翻录语音均能被 100% 识别；当回放设备为 Yamaha TSX-140 时，偷录设备为 Sony PX440 和 iPhone6 的翻录语音检测率为 100%，偷录设备为 Mi4 的翻录语音检测率达到了 99.71%；当回放设备为 Philips DTM3115 时，识别准确率虽不及以上两种，但其准确率也达到了 99.7% 以上。实验结果表明：本文检测算法能够很好地检测翻录语音与原始语音，且对偷录设备与回放设备有较好的顽健性。

5.3 加载翻录语音检测模块后的识别系统

为更好地检验翻录语音对说话人识别系统的攻击情况，本文分别在 GMM-UBM 和 i-vector 说话人识别系统上进行了实验。在实验中，提取 13 维 MFCC 基本特征，与一阶、二阶差分构成 39 维特征参数。翻录语音检测模块使用 FFT 点数为 1 024。实验中训练了 4 个用户模型，在测试时用每个用户的翻录语音作为攻击语音，其中每个用户模型使用原始语音 130 个、翻录语音 150 个。

科研工作者从 20 世纪 50 年代开始进行说话人识别技术研究，至今取得了一定的进展^[16]。GMM-UBM 和 i-vector 说话人识别系统是经典和目前主流的两个系统。GMM-UBM^[17]主要利用 UBM 和少量的说话人数据，通过自适应算法得到目标人模型，最后用测试数据分别与模型和 UBM

进行打分比较；基于 i-vector 并进行信道补偿的说话人识别系统^[18]根据 UBM 和 T 子空间提取 i-vector，建立 GPLDA (Gaussian probabilistic linear discriminate analysis，高斯概率线性判别分析) 模型及说话人模型，测试阶段用测试语音对模型进行打分比较，本实验打分方式采用对数似然比。

本实验分别在 GMM-UBM 和 i-vector 说话人识别系统上，测试来源不同的回放设备和偷录设备的翻录语音是否能够攻击成功，检测结果如图 10 (a) 和图 10 (b) 所示。在 GMM-UBM 系统上，未加载翻录语音检测模块的说话人识别系统的 EER 为 46.33%，加载了翻录语音检测模块后，该系统的 EER 降为 6.48%，下降了约 40%。在 i-vector 系统上，系统对翻录语音的攻击有一定的防御性，但并不能完全抵抗攻击，该系统的 EER 为 27.78%，当加载了翻录语音检测模块时，系统 EER 降为 5.56%，下降了约 22%。该数据说明本检测算法能够有效地提高 GMM-UBM 和 i-vector 识别系统抵抗翻录语音攻击的能力。

5.4 对比实验

将本文算法与典型的 3 种算法在本文数据库上进行对比，实验结果见表 7。参考文献[5]算法采用短时能量法提取静音，利用谱减法进行滤波，提取 MFCC 特征参数。参考文献[7]方法采用高通滤波器进行去噪，提取信道模式噪声，并提取 6 个统计特征及 6 阶 Legendre 多项式系数。参考文献[10]中算法将语音信号进行 MP3 编码后，提取比例因子统计特征作为检测特征。表 7 中，ACC 表示准确率，EER 表示将检测模块加载到 GMM-UBM、i-vector 说话人识别系统时，系统的等错误概率。

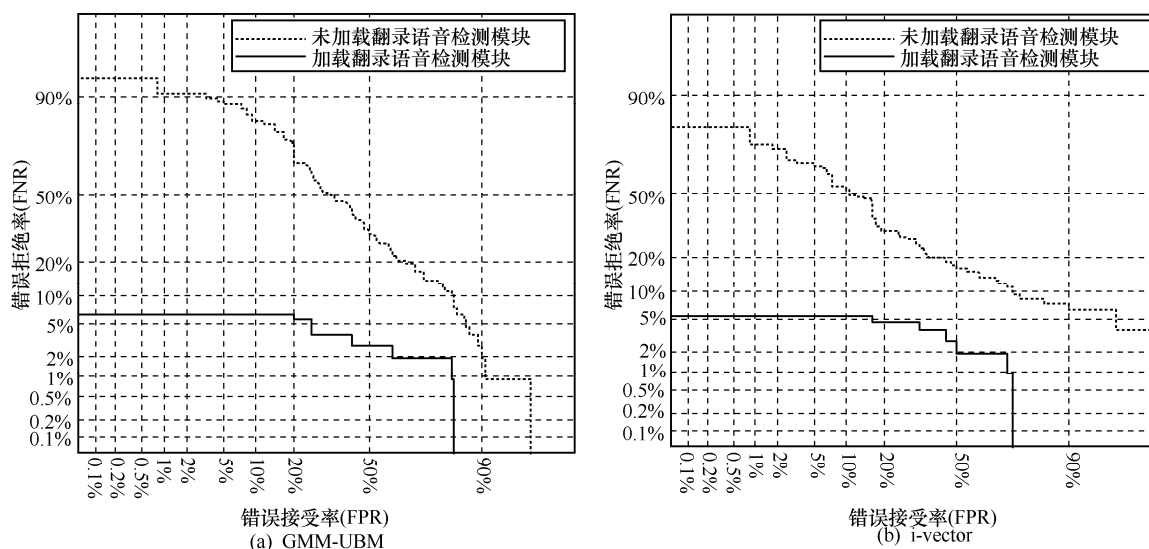


图10 加载翻录语音检测模块前后的等错误概率对比

表7 本文算法与参考文献[5,7,10]算法检测结果的对比

算法	ACC	EER	
		GMM-UBM	i-vector
本文算法	99.04%	6.48%	5.56%
参考文献[5]算法	73.70%	26.67%	18.51%
参考文献[7]算法	78.28%	23.33%	17.50%
参考文献[10]算法	98.91%	7.03%	6.40%

由表7可以看出,对于多种偷录及回放设备的翻录语音,本文算法优于参考文献[5,7,10]算法,识别率分别提高了约26%、约21%和约0.1%。解决了参考文献[5,7]中设备过于单一的问题,更具实用性。另外,参考文献[5,7,10]中算法使用特征维数分别为39维、12维和21维,本文算法共计10维特征,维数更低。

6 结束语

本文针对多种偷录设备与回放设备的翻录语音攻击,提出了基于相位谱的翻录语音检测算法。并通过模拟实际翻录语音攻击的整个物理过程,建立了实验语音数据库。本文确定了最佳检测效果的FFT点数,对来源不同录音设备的翻录语音进行了检测,其检测率达到了99.04%,在此基础上,将该检测算法模块加载到说话人识别系统中,其抵抗翻录语音攻击的能力提升了约22%。在今

后的研究中,将进一步探究各种录音设备及回放设备对语音的影响,并且在检测方法上进行创新和改进。

参考文献:

- [1] SHANG W, STEVENSON M. A playback attack detector for speaker verification systems[C]//2008 IEEE International Symposium on Communications Control and Signal Processing (ISCCSP), March 12-14, 2008, Bordeaux, France. New Jersey: IEEE Press, 2008:1144-1149.
- [2] SHANG W, STEVENSON M. Score normalization in playback attack detection[C]//IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), March 14-19, 2008, Dallas, USA. New Jersey: IEEE Press, 2010:1678-1681.
- [3] JAKUB G, MARCIN G, RAFAL S. Playback attack detection for text-dependent speaker verification over telephone channels [J]. Speech Communication, 2015(67):143-153.
- [4] WU Z, GAO S, CLING E S, et al. A study on replay attack and anti-spoofing for text-dependent speaker verification[C]//IEEE 2014 Summit and Conference, Asia-Pacific Signal and Information Processing Association, December 9-12, 2014, Siem Reap, Cambodia. New Jersey: IEEE Press, 2014: 35-45.
- [5] 张利鹏, 曹攀, 徐明星. 防止假冒者闯入说话人识别系统[J]. 清华大学学报(自然科学版), 2008, 48(S1): 699-703.
ZHANG L P, CAO J, XU M X. Prevention of impostors entering speaker recognition systems[J]. Journal of Tsinghua university (Science and Technology), 2008, 48(S1): 699-703.
- [6] 王志锋, 贺前华, 张雪源, 等. 基于模式噪声的录音回放攻击检测[J]. 华南理工大学学报, 2011, 39(10): 7-12.
WANG Z F, HE Q H, ZHANG X Y, et al. Channel pattern noise based playback detection algorithm speaker recognition[J]. Journal of South China University of Technology(Natural Science Edition), 2011, 39(10): 7-12.



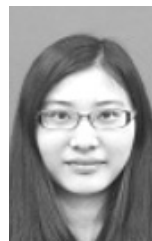
- [7] WANG Z F, HE Q H, ZHANG X Y, et al. Channel pattern noise based playback detection algorithm speaker recognition [C]// IEEE International Conference on Machine Learning and Cybernetics(ICMLC), July 10-13, 2011, Guilin, China. New Jersey: IEEE Press, 2011: 1708-1713.
- [8] VILLABA J, LLEIDA E. Detecting replay attacks from far-field recordings on speaker verification systems[C]//COST 2011 European Conference on Biometrics and ID Management, March 8-10, 2011, Brandenburg, Germany. New York: ACM Press, 2011: 274-285.
- [9] VILLABA J, LLEIDA E. Preventing replay attacks on speaker verification systems[C]//IEEE International Carnahan Conference on Security Technology (ICCST), October 18-21, 2011, San Francisco, USA. New Jersey: IEEE Press, 2011: 1-8.
- [10] CHEN Y N, WANG R D, YAN D Q, et al. Voice playback detection based on long-window scale-factors[J]. International Journal of Security and Its Application, 2016, 10(12): 299-310.
- [11] 郑志彬. 信息网络安全威胁及技术发展趋势[J]. 电信科学, 2009, 25(2): 28-34.
- ZHENG Z B. Overview of mobile communication services security[J]. Telecommunications Science, 2009, 25(2): 28-34.
- [12] 王帅, 汪来富, 金华敏, 等. 网络安全分析中的大数据技术应用[J]. 电信科学, 2015, 31(7): 145-150.
- WANG S, WANG L F, JIN H M, et al. Big data application in network security analysis [J]. Telecommunications Science, 2015, 31(7): 145-150.
- [13] OPPENHERIM A V, LIM J S. The important of phase in signals[J]. Processing of the IEEE, 1981, 69(5): 529-541.
- [14] DUAN K B, RAJAPAKSE J C, WANG H Y, et al. Multiple SVM-RFE for gene selection in cancer classification with expression data[J]. IEEE Transactions on Nano Bioscience, 2005, 4(3): 228-234.
- [15] 王天庆, 李爱军. 连续汉语语音识别语料库的设计[C]// 第六届全国现代语音学学术会议论文集(下), 2003 年 10 月 18-20 日, 天津, 中国. 天津: 天津人民出版社, 2003.
- WANG T Q, LI A J. The design of the continuous Chinese speech recognition corpus[C]//The sixth national conference on modern phonetics learning, October 18-20, 2003, Tianjin, China, Tianjin: Tianjin Remin Chubanshe, 2003.
- [16] 杨震, 徐敏捷, 刘璋峰, 等. 语音大数据信息处理架构及关键技术研究[J]. 电信科学, 2013, 29(11): 1-5.
- YANG Z, XU M J, LIU Z F, et al. Study of audio frequency big data processing architecture and key technology[J]. Telecom

munications Science, 2013, 29(11): 1-5.

- [17] CHAKROBORTY S, ROY A, SAHA G. Improved closed set text-independent speaker identification by combining MFCC with evidence from flipped filter banks[J]. International Journal of Signal Processing, 2007, 4(2): 114-122.

- [18] KANAGASUNDARAM A, DEANA D, SRIDHARAN S, et al. I-vector based speaker recognition using advanced channel compensation techniques[J]. Computer Speech and Language, 2014, 28(1): 121-140.

[作者简介]



李臻(1992-), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为多媒体通信与信息安全等。



王让定(1962-), 男, 博士, 宁波大学高等技术研究院教授、博士生导师, 主要研究方向为多媒体通信与取证、信息隐藏与隐写分析、智能抄表及传感网络技术等。



严迪群(1979-), 男, 博士, 宁波大学信息科学与工程学院副教授、硕士生导师, 主要研究方向为多媒体通信、信息安全、基于深度学习的数字语音取证等。



陈亚楠(1990-), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为多媒体通信与信息安全等。